OMICS International conferenceseries.com

2ⁿᵈ International Conference on
Big Data Analysis and Data Mining
November 30-December 01, 2015   San Antonio, USA

Enhancement of enterprise search with neural language model

Pengchu Zhang, Laritza Saenz and **John Mareda**
Sandia National Laboratories, USA

A significant problem that reduces the effectiveness of enterprise search is query terms that do not exist in the enterprise data. Consequently, enterprise search generates no results or the answers match the exact query terms and do not take into account related terms. This results in a high rate of false positives in terms of information relevance. Recent developments in neural language model (NLM), specifically, the word2vec model initiated by Google researchers has drawn a great deal of attention in last two years. This model uses multiple layers of neural networks to represent words into vector spaces. The vector representation of words carries both semantic as well syntactic meanings. Terms with the semantic similarities are close together in the vector space as measured by their Euclidean distances. Enterprise search may utilize the "contextual" relationships between words to intelligently increase the breath and quality of search results. Application of the NML in our enterprise search promises to significantly improve the findability and relevance of returned information. We expand the query term(s) into a set of related terms using the trained term vectors based on corporate data repositories as well as making use of Wikipedia. The expanded set of terms is used to search the indexed enterprise data. The most relevant data rises in ranking including documents which may not contain the original query terms. In this presentation, we will also discuss the potential and limitations of applying NLM in search and other aspects of enterprise knowledge management.

Biography

Pengchu Zhang has more than 10 years experiences in computer modeling/simulation, machine learning, data mining and unstructured data analysis at Sandia National Laboratories. His recent research interest is to develop and apply technologies of deep learning in enterprise knowledge sharing and management.

pzhang@sandia.gov

Notes:

J Data Mining Genomics Proteomics
ISSN: 2153-0602 JDMGP, an open access journal

Data Mining 2015
November 30-December 01, 2015

Volume 6 Issue 4

Page 25